

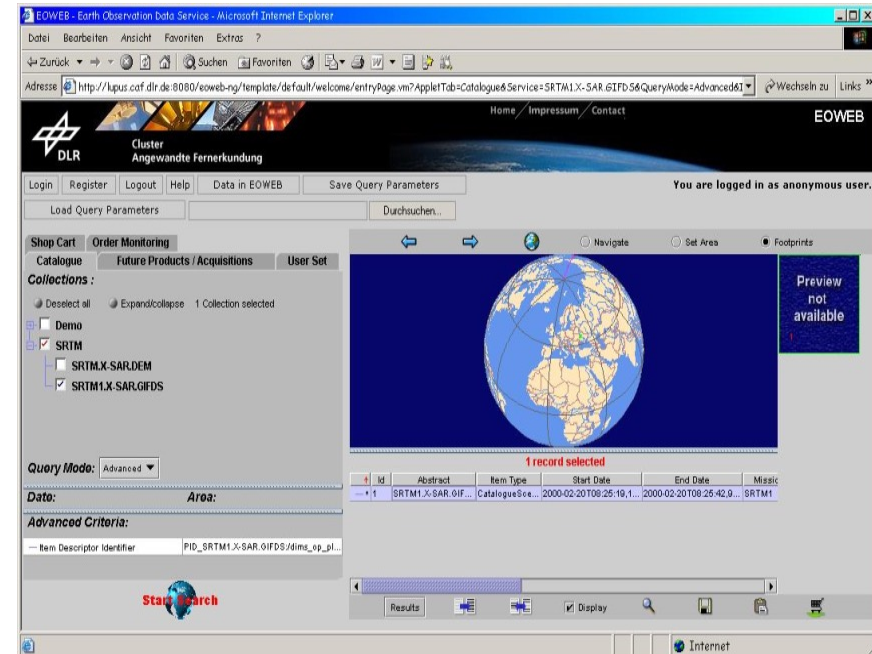
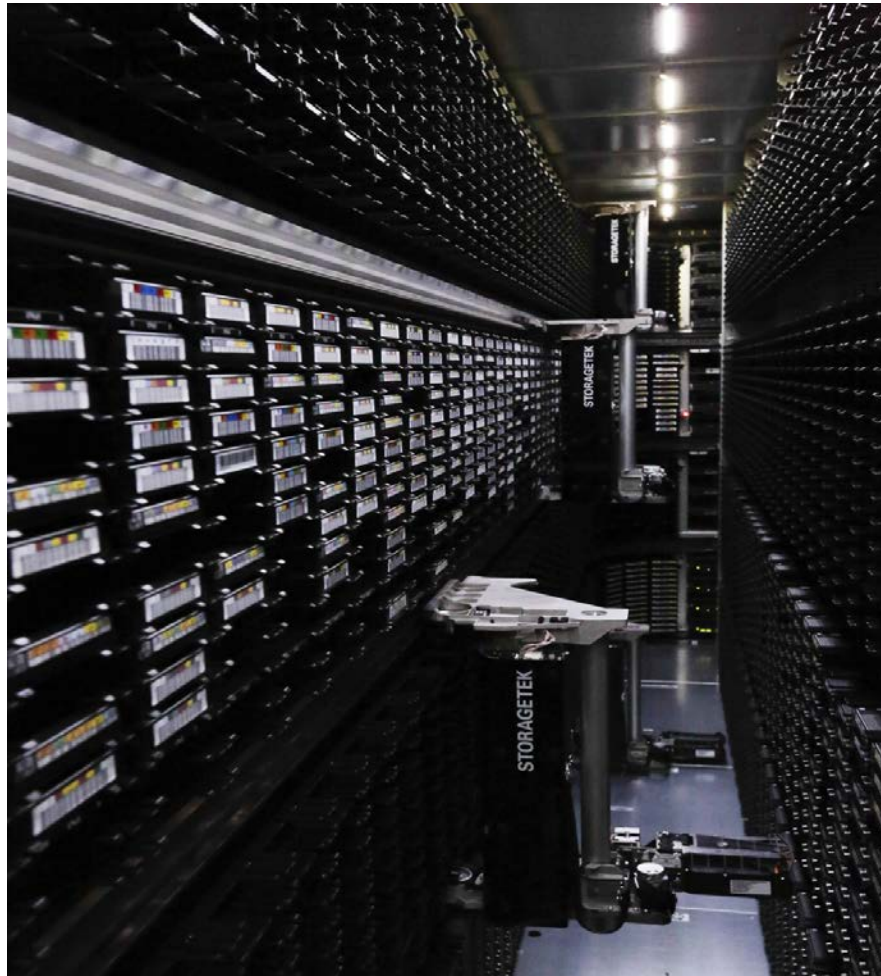
Big Data Analytics for Detailed Urban Mapping

Mihai Datcu

Daniela Molina Espinoza, Octavian Dumitru, Gottfried Schwarz



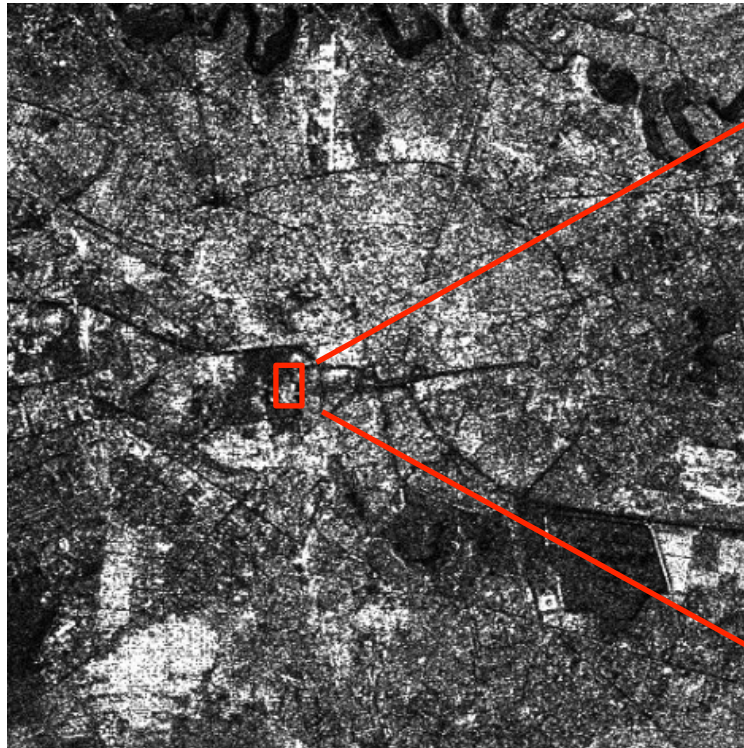
Big Data: The German EO Digital Library



The data access



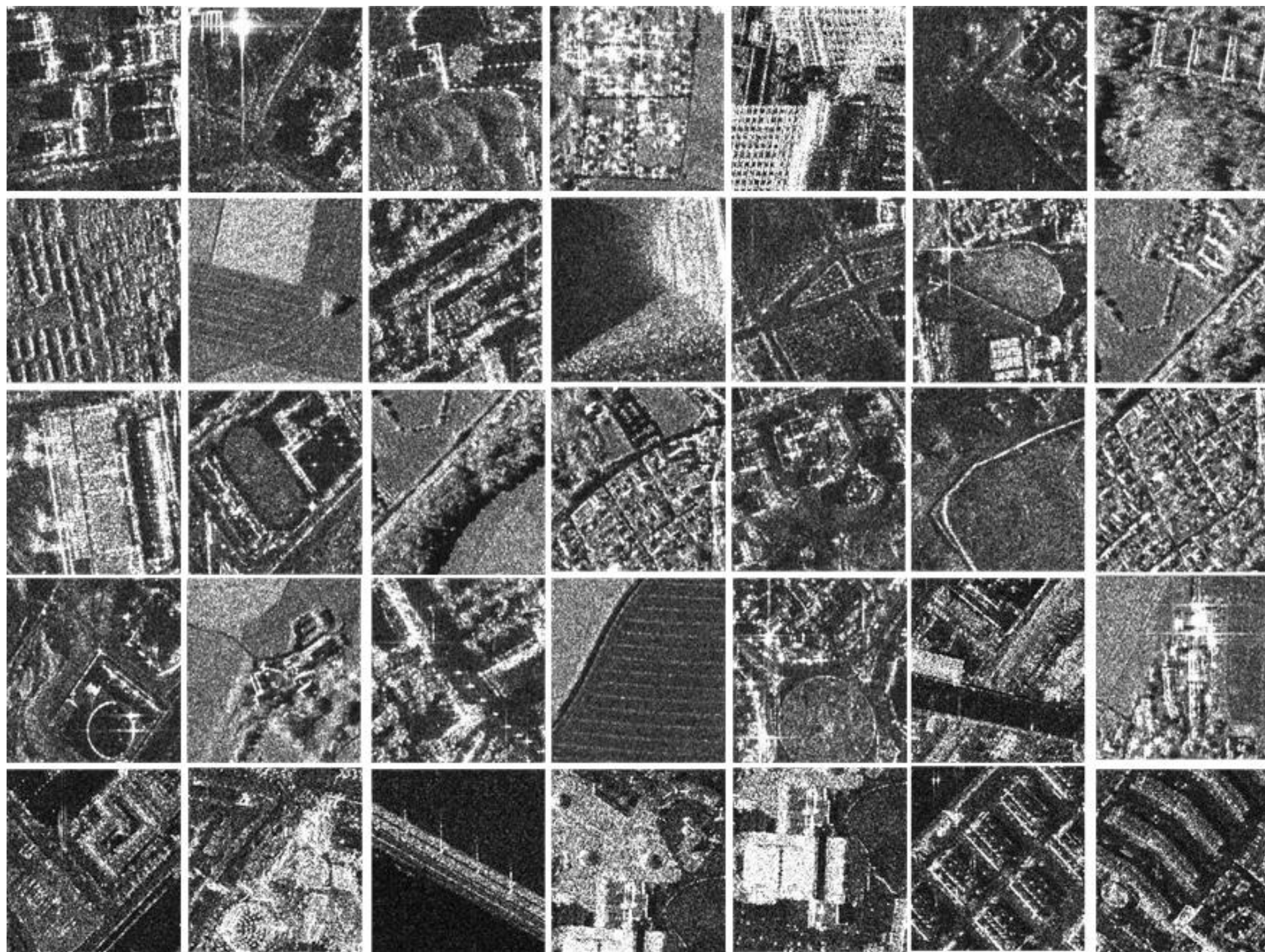
Information vs. Data

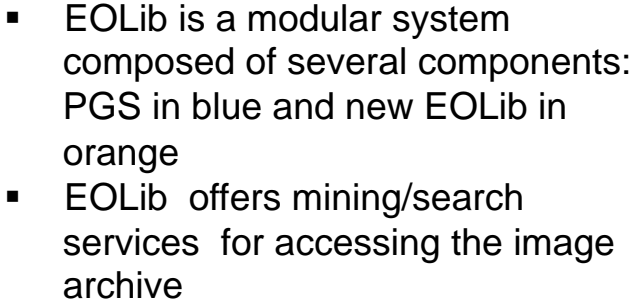


ERS1
24-JUL-1992
512x512 pixels



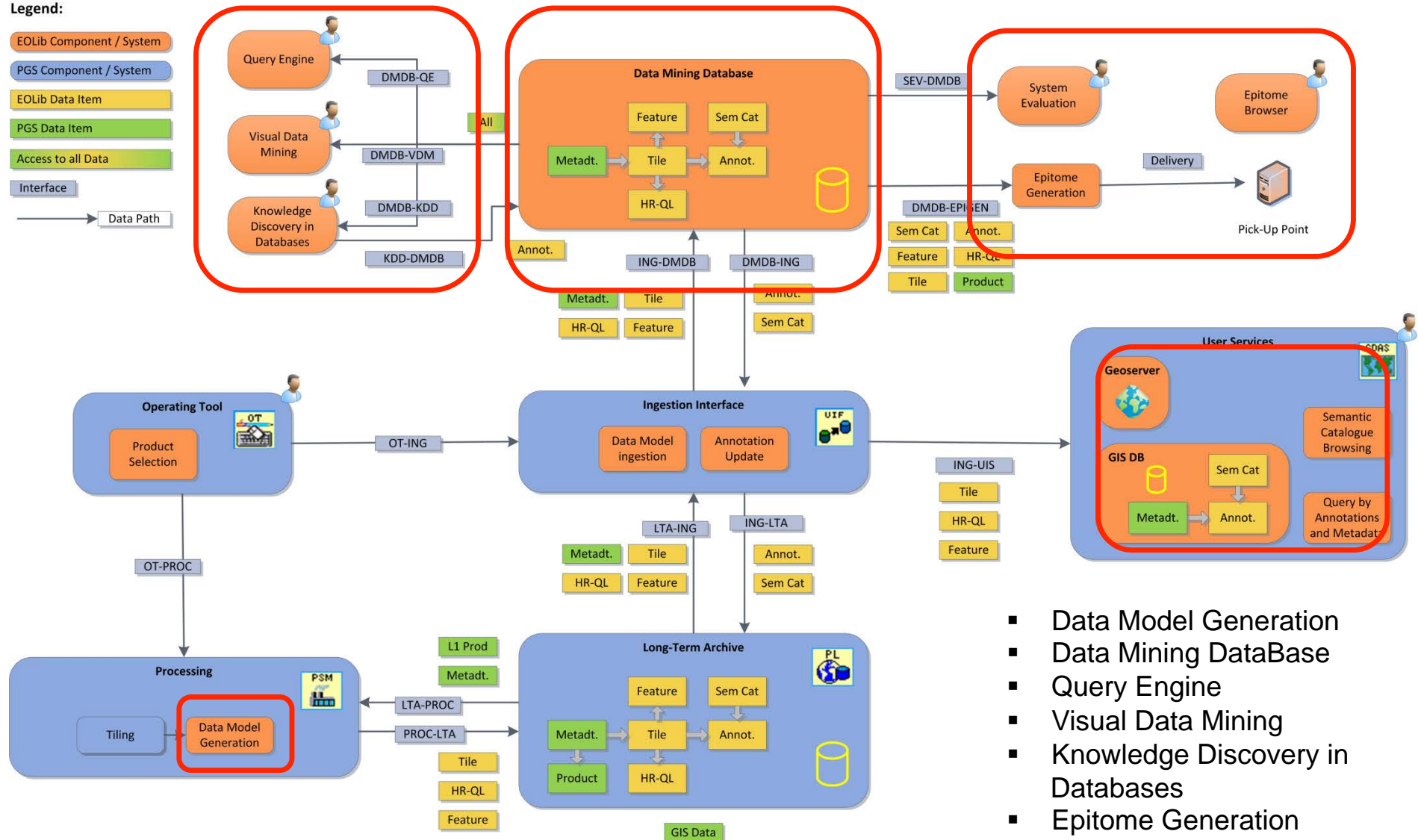
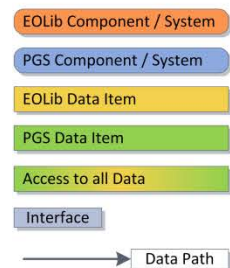
TerraSAR-X
11-OCT-2008
512x512 pixels





EOLib: Data Mining and KDD new components

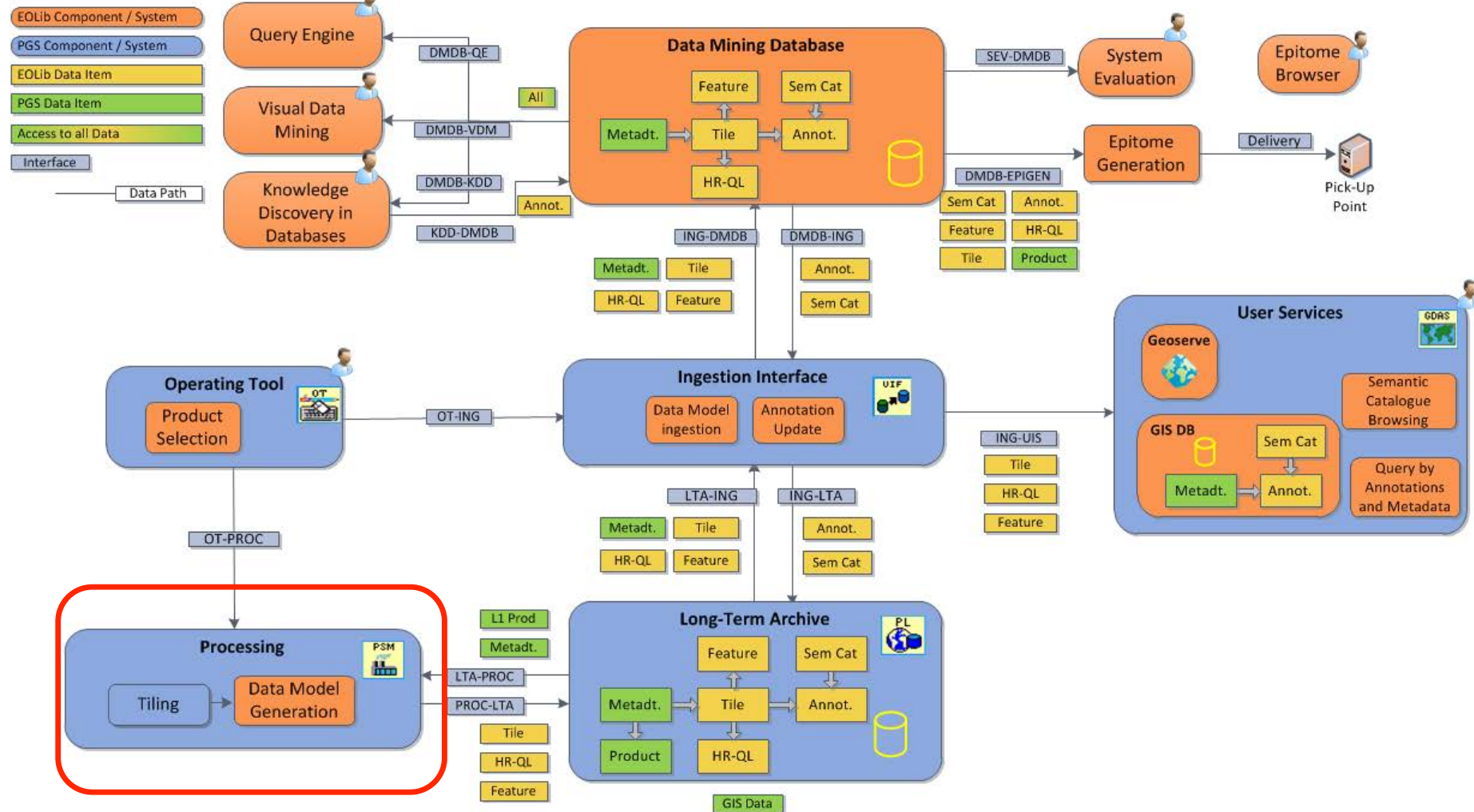
Legend:



- Data Model Generation
- Data Mining DataBase
- Query Engine
- Visual Data Mining
- Knowledge Discovery in Databases
- Epitome Generation

Data Model Generation

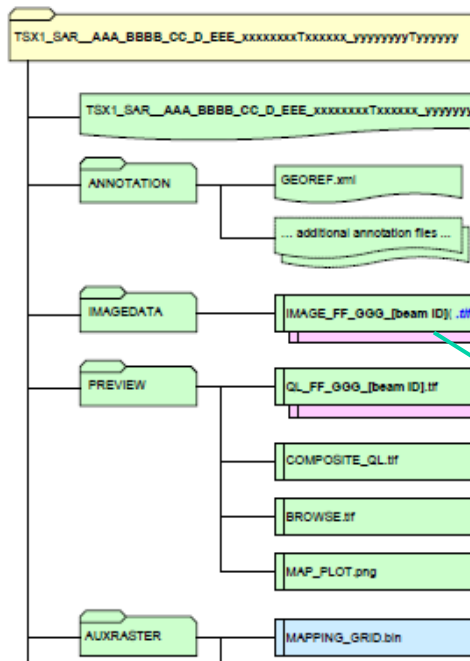
Legend:





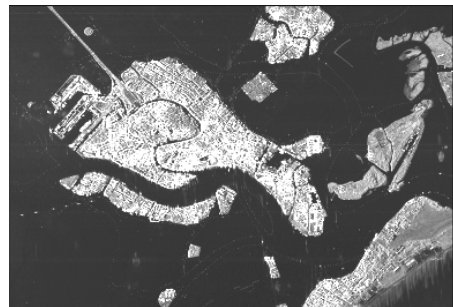
Data Model Generation

TerraSAR-X L1b product

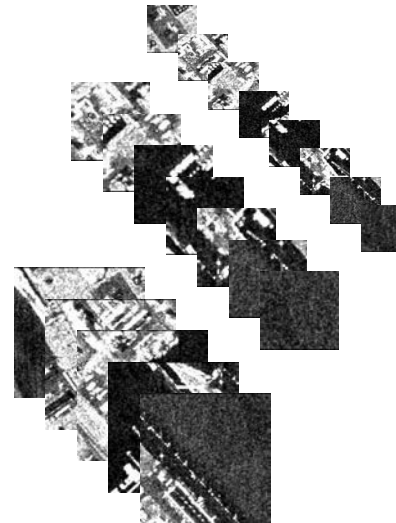


TerraSAR-X image

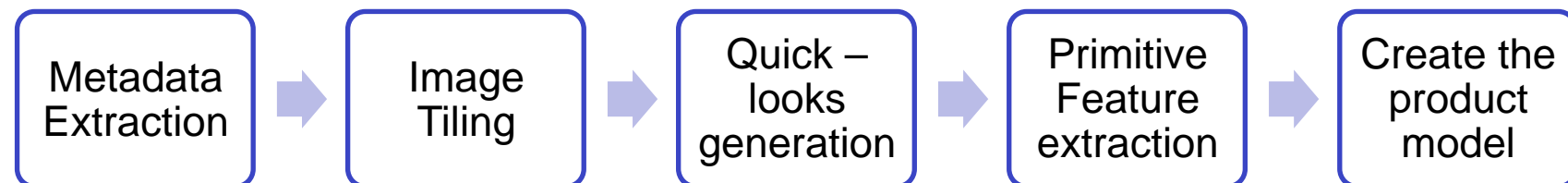
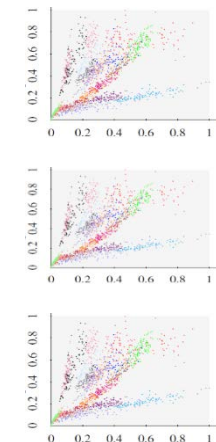
```
<productInfo>
  <missionInfo>
    <mission>TSX-1</mission>
    ...
  </missionInfo>
  <acquisitionInfo>
    ...
  </acquisitionInfo>
  ...
</productInfo>
```



Tiles with different size



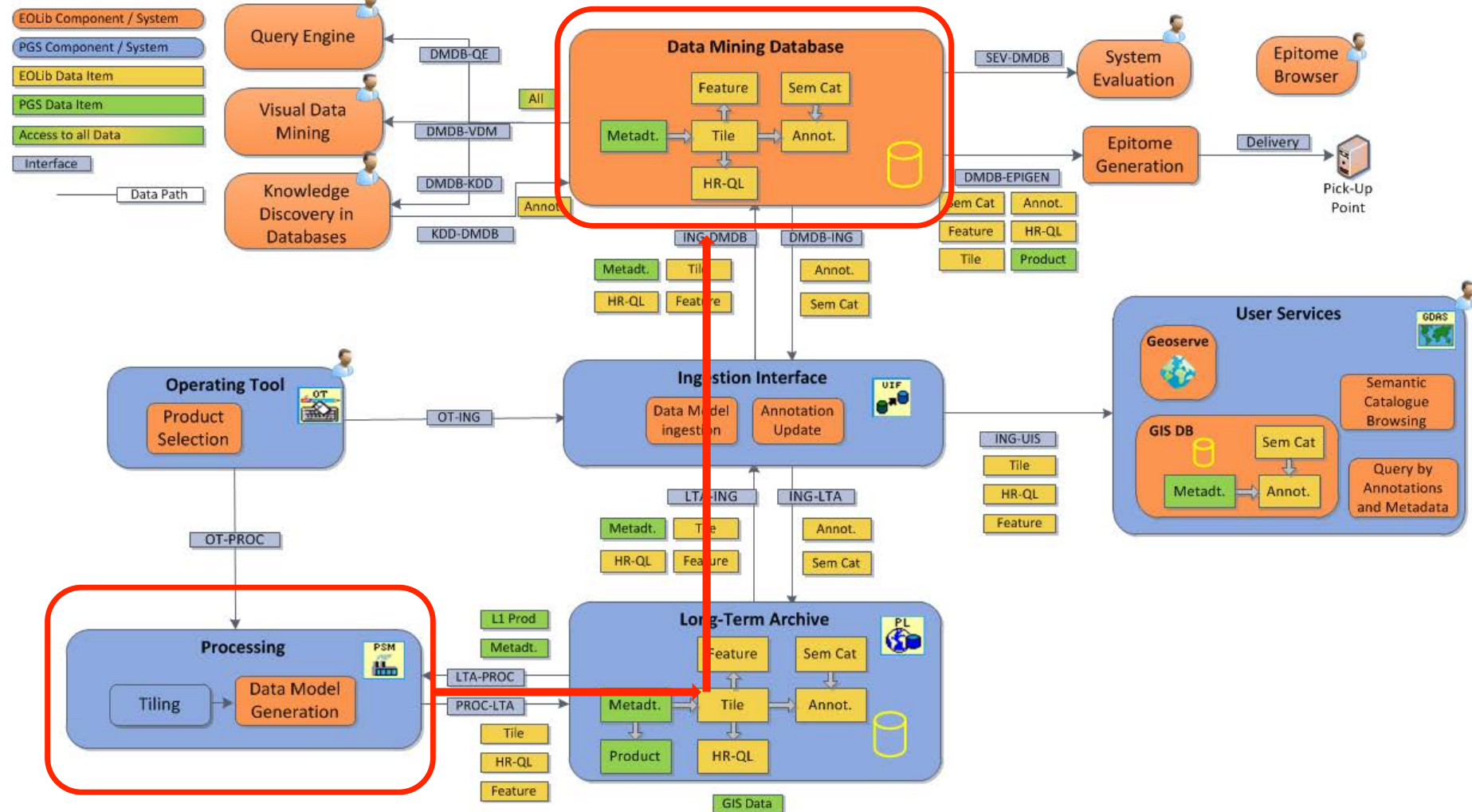
Primitive features: Gabor filter and weber local descriptor



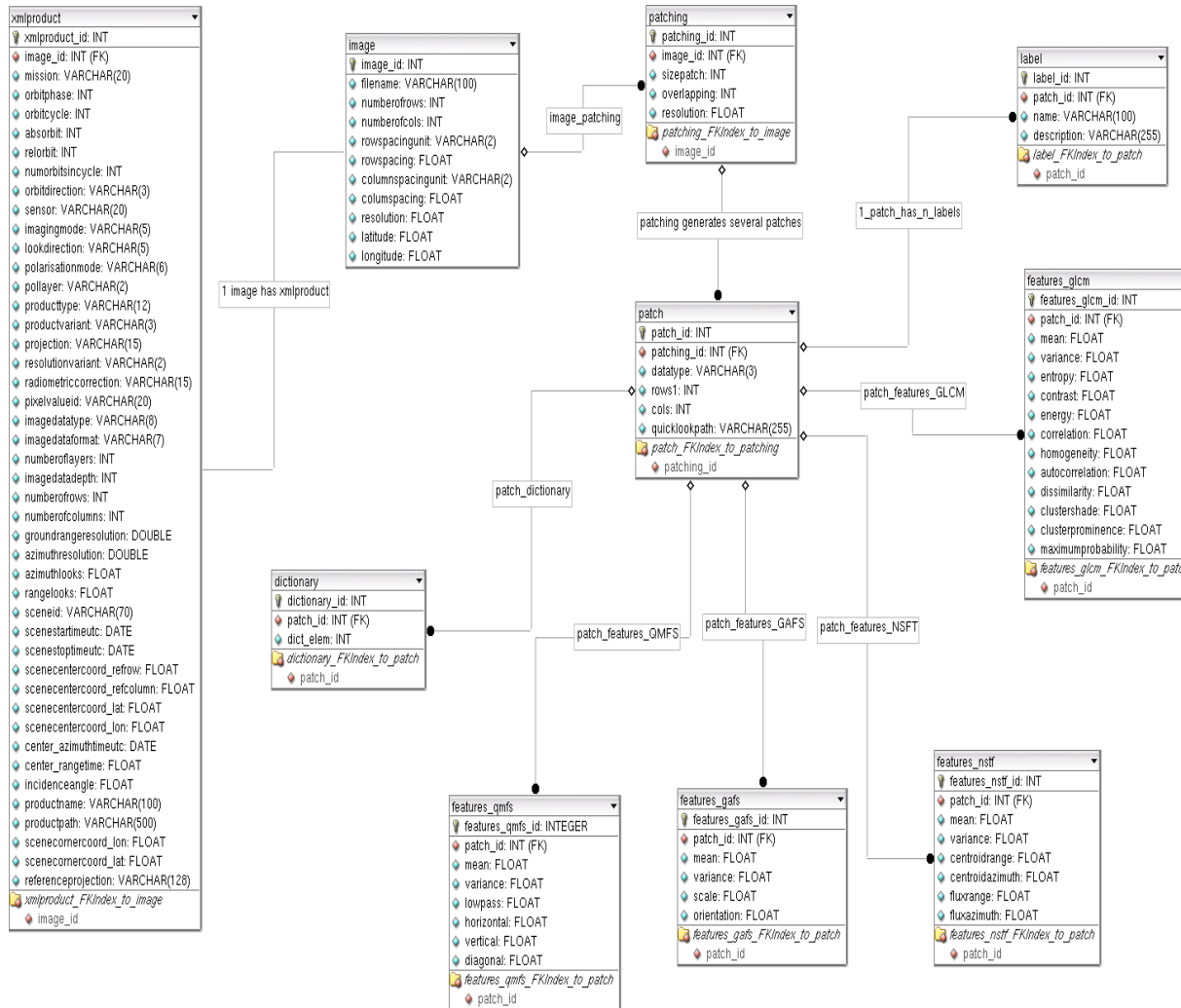


Data Mining Database

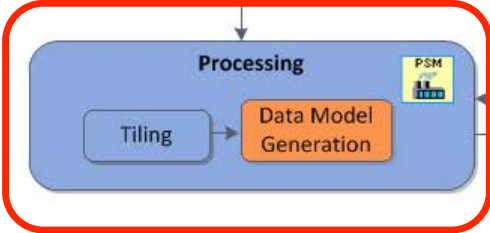
Legend:



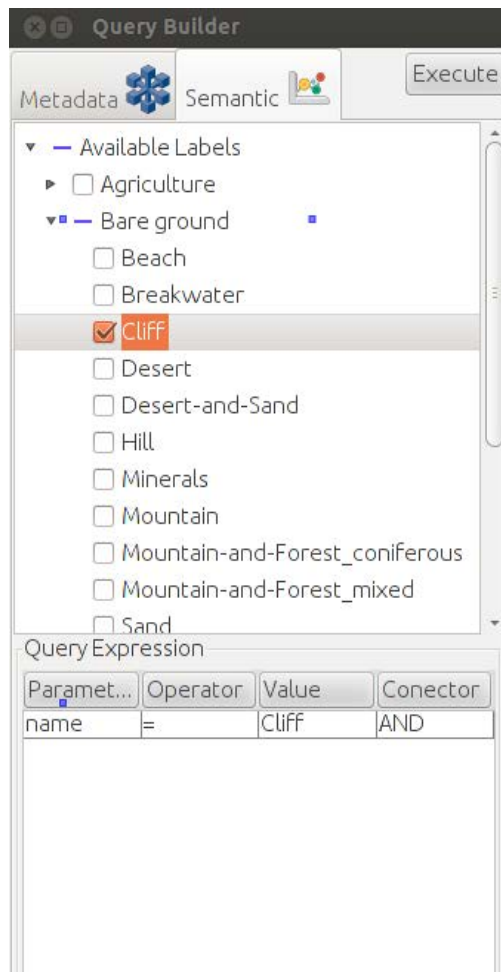
Data Mining Data Base



- It is a relational database
- DMDB comprises about
 - 800 processed products
 - 8. millions of tiles
 - 20 thousand metadata entries.
 - 106 semantic labels



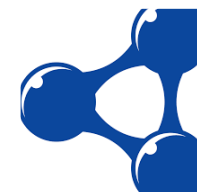
Query Engine



Metadata

- Coordinates (lat/lon)
- Incidence angles
- Acquisition time
- Pixel spacing
- Number of columns/ rows
- sensor
- Mission
- orbits

Metadata parameter based on XML annotation file of TerraSAR-X L1b products



Semantics

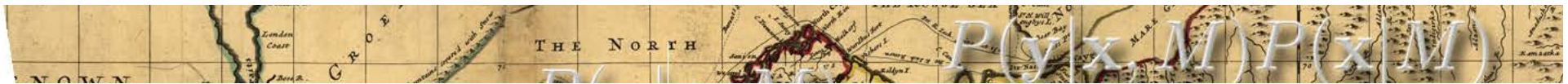
- **Agriculture**
 - Cropland
 - Rice plantation.....
- **Bare ground**
 - Cliff
 - Desert.....
- **Forest**
 - Forest coniferous
 - Forest mixed....
- **Urban area**
 - Commercial areas
 - High density residential areas....

Semantic parameters based on EO Taxonomy

Query Engine: Examples

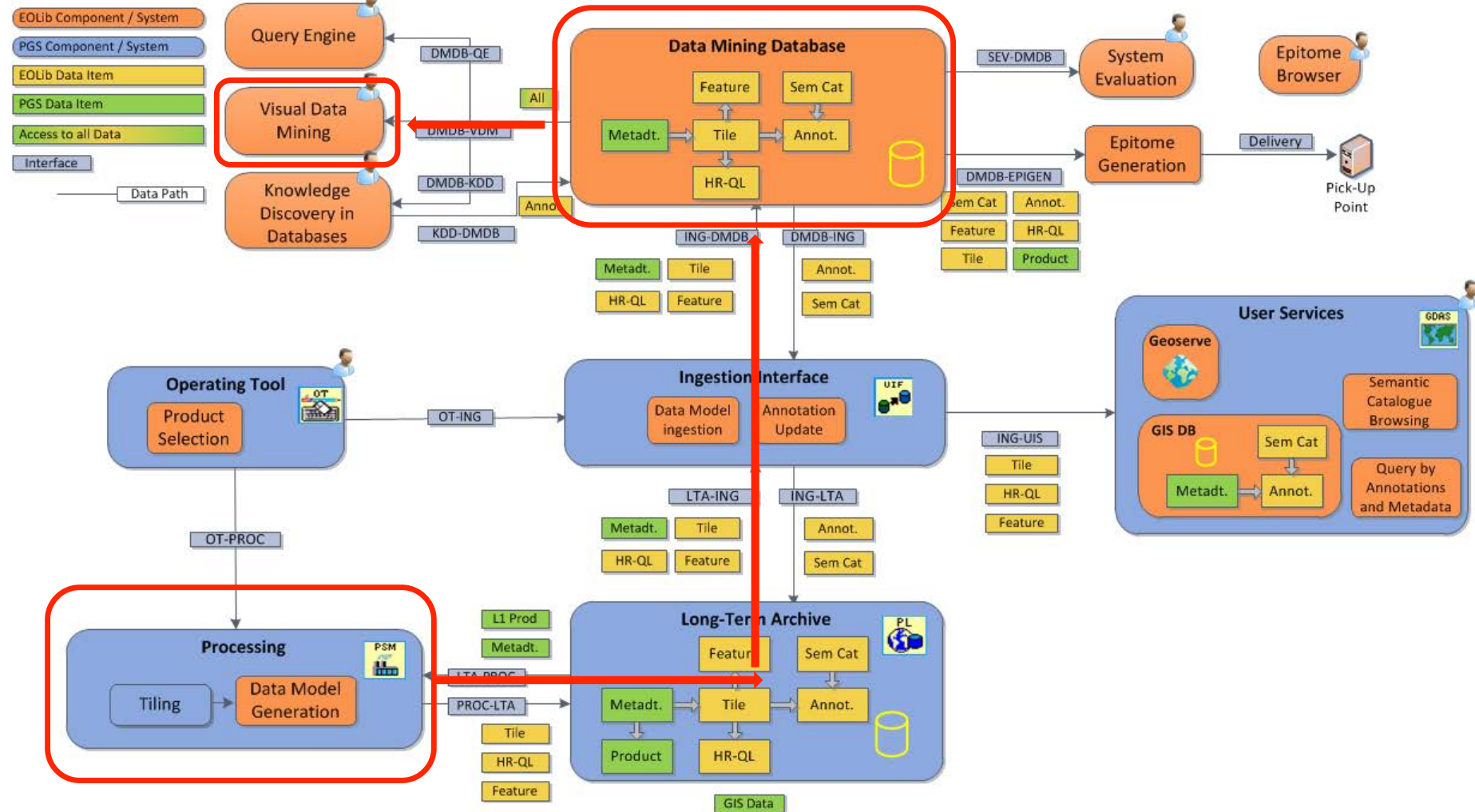
The screenshot displays the Query Engine interface with two main panels. The left panel, titled '(on dolphin)', shows a tree view of metadata. Under the 'Semantic' tab, 'Storage tanks' and 'Medium density residential area' are selected, indicated by red boxes. The 'Query Expression' section at the bottom shows a query built with 'name = Storage tanks OR name = Medium density residential area'. The right panel, titled 'Query Results.. (on dolphin)', shows a table of results with columns: tile_id, label_id, goodness, coverage, trust, lastupdate, label_id, parentlabel, name, description, level, source, and stat. The table contains 20 rows of data. Below the table is a grid of 20 small satellite images, each corresponding to a row in the table. A red box highlights the image for the row where label_id is 435520, which is labeled '435520 Storage tanks'.

tile_id	label_id	goodness	coverage	trust	lastupdate	label_id	parentlabel	name	description	level	source	stat
436990	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435010	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435519	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436737	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435254	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
433506	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436410	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435520	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436045	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436823	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436371	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436132	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436214	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436667	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435897	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435589	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436199	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436734	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436828	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436416	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
436412	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435423	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435546	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1
435459	55	0.0	0.0	0.0	2015-10...	55	3	Storage ta...	Descriptio...	2	EOT	1



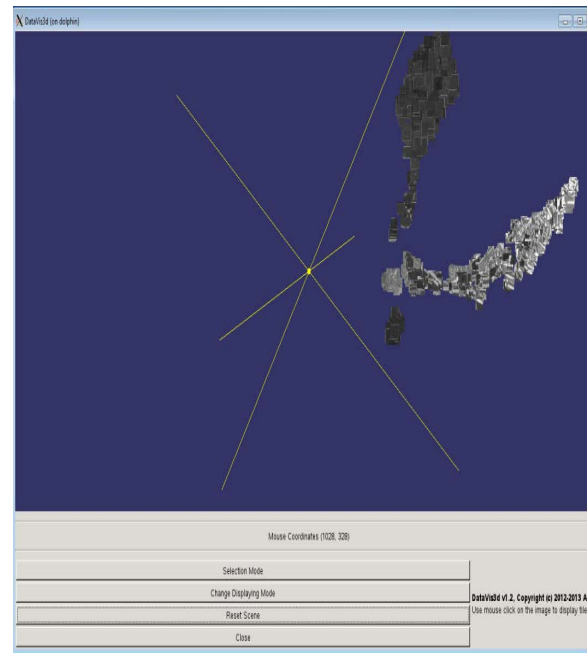
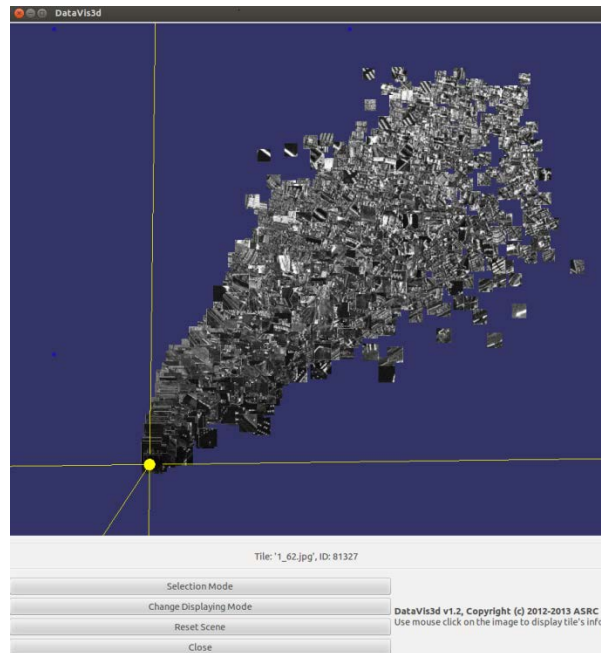
Visual Data Mining

Legend:

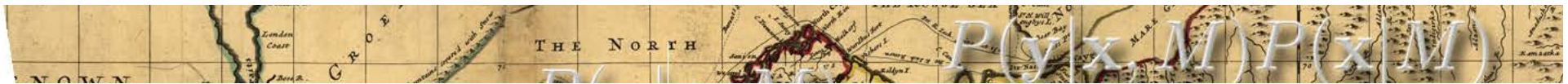




Visual Data Mining

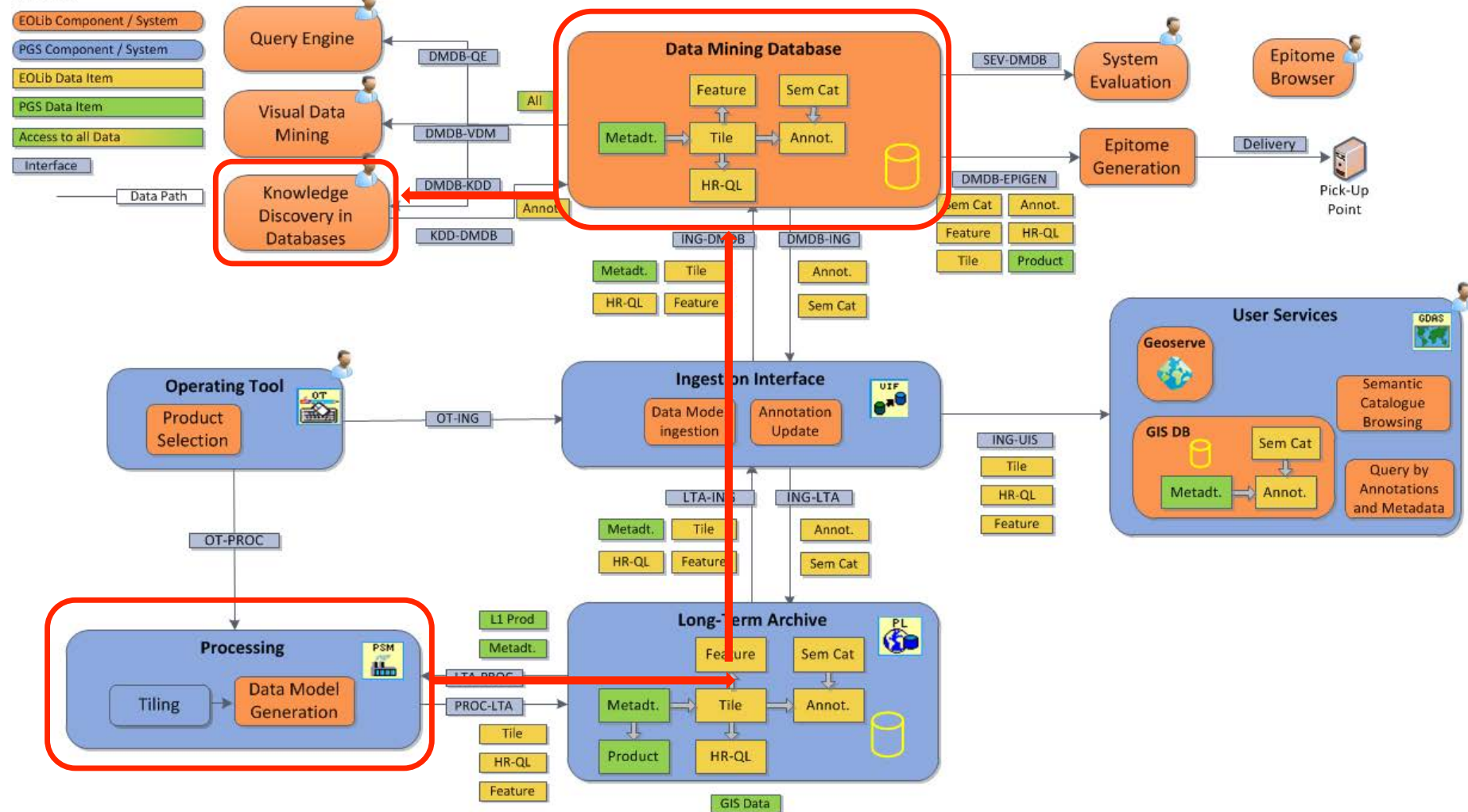


- Provides a projection of the entire database
- Representation of the data in the 3D space (dimensionality reduction)
- Interactive exploration and analysis of very large, high complexity data sets
- This allows the user:
 - To browse the image archive
 - To find scenes of interest
- Semantically consistent groups may appear inside the data

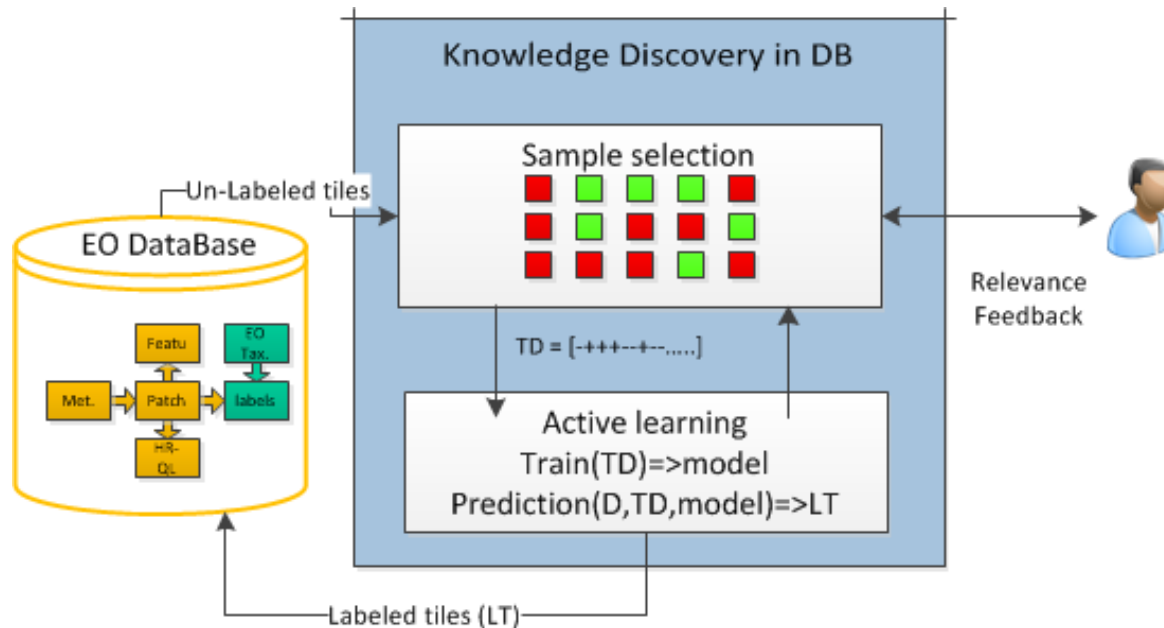


Knowledge Discovery in Databases

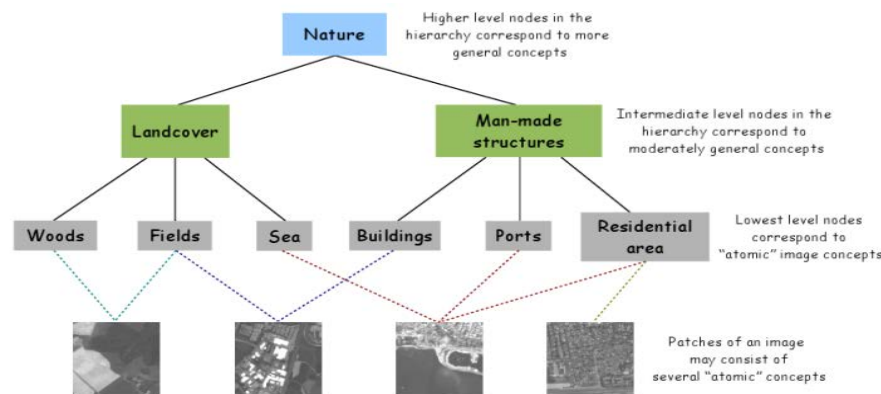
Legend:



Knowledge Discovery in Databases

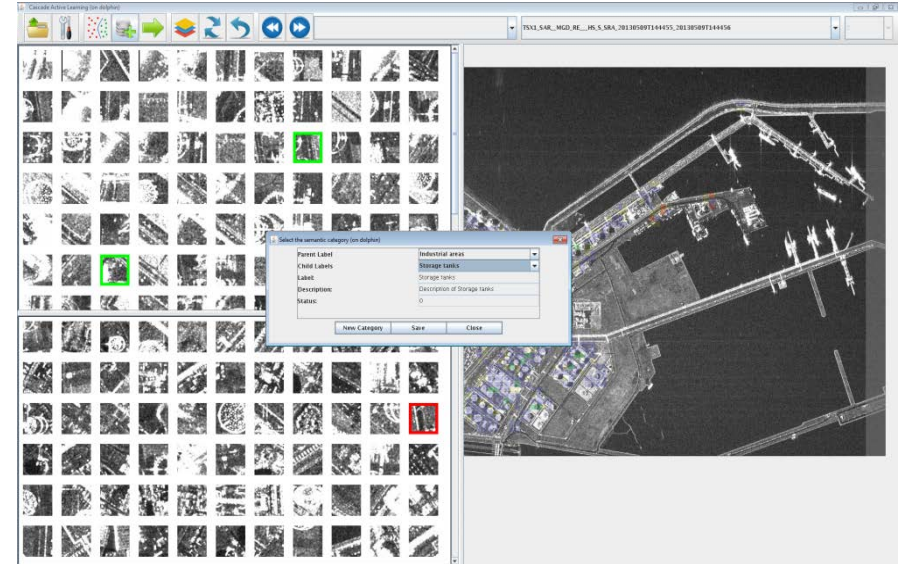
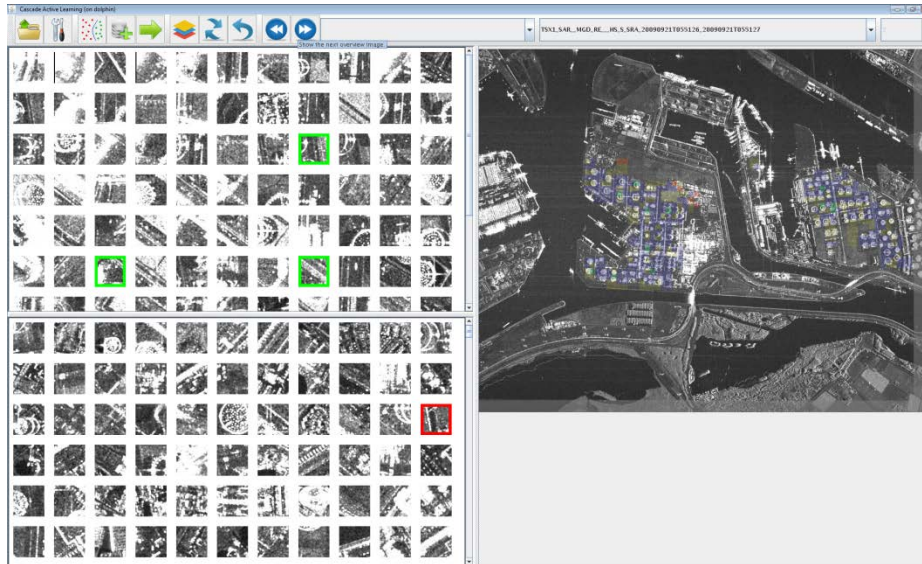


- KDD used to define **semantic annotations of the image content**.
- Interactive search supported by relevance feedback mechanisms
- Goals is to build a model which performs the mapping between low-level image descriptors (primitive features) and high-level image concepts (semantics)

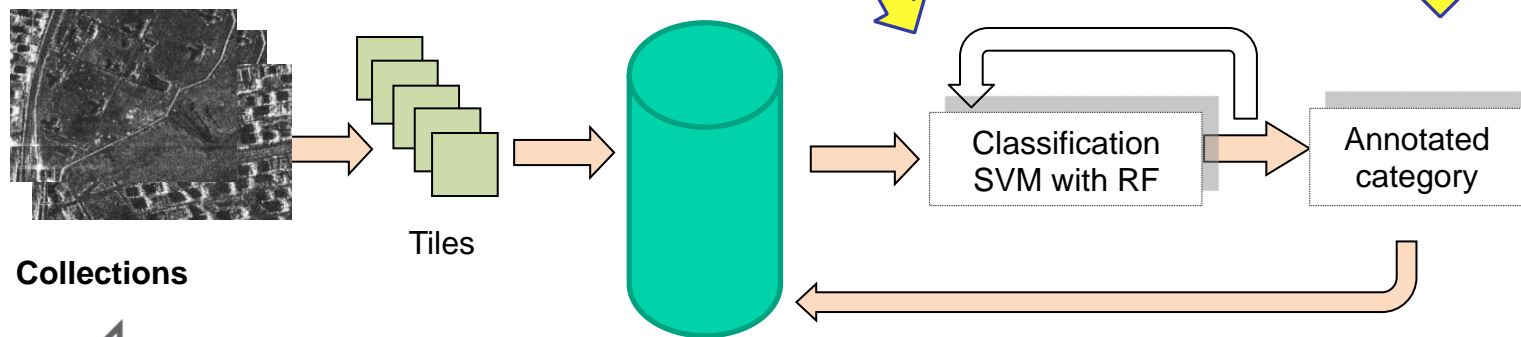




KDD: GUI



Work Flow:

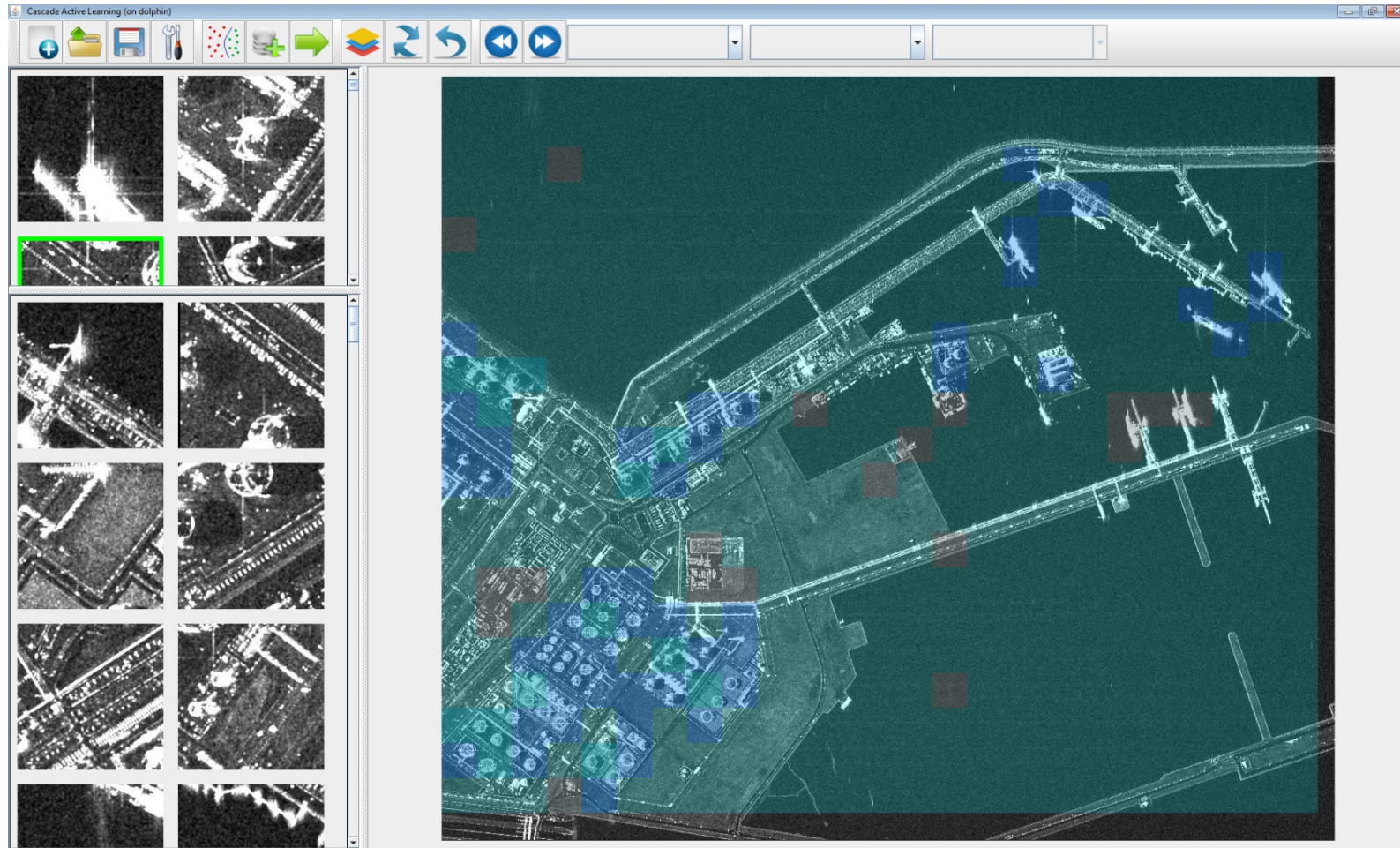






Cascaded Active Learning

Level 0: 200x200 pixels



Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft



Cascaded Active Learning

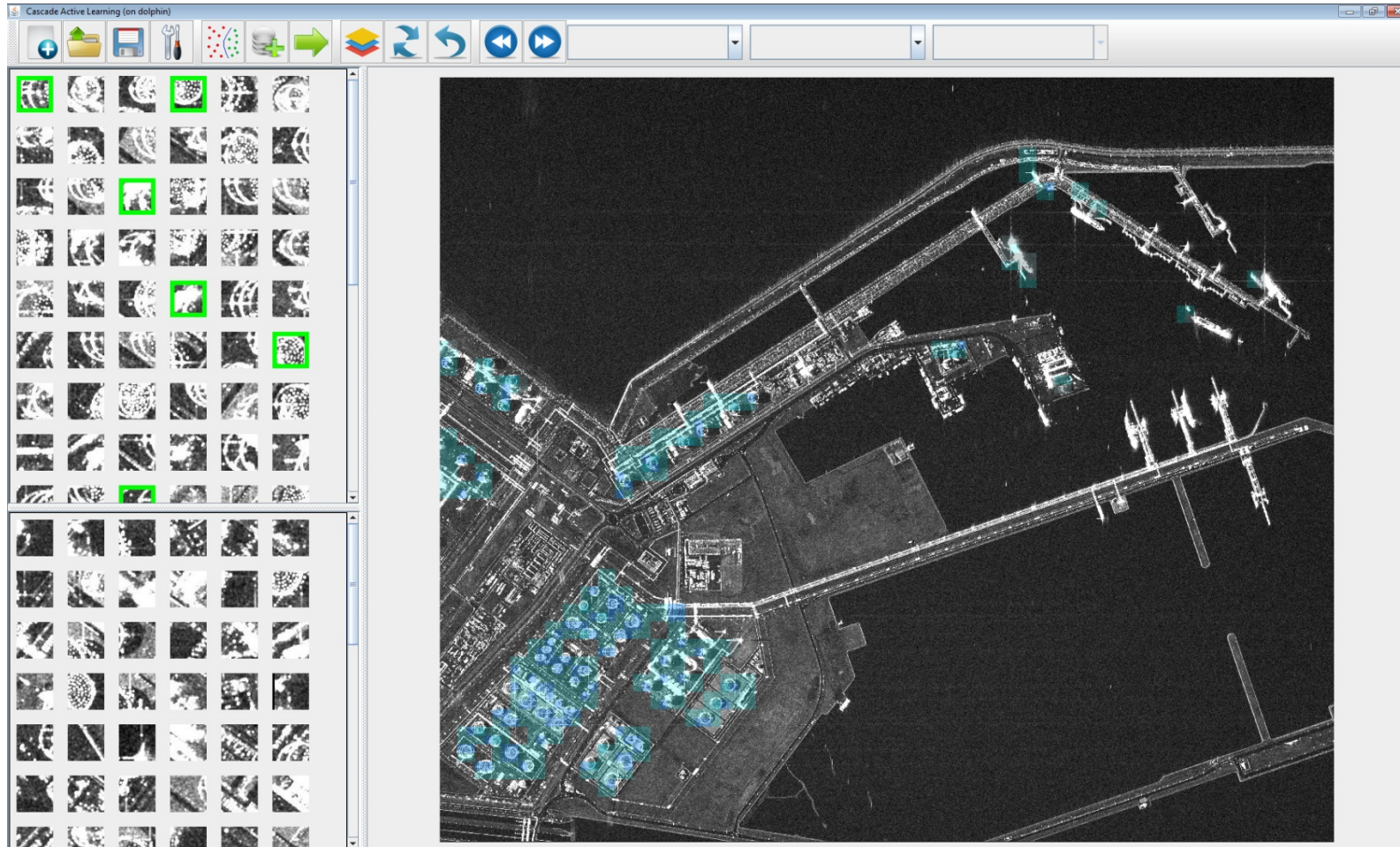
Level 1: 100x100 pixels





Cascaded Active Learning

Level 2: 50x50 pixels

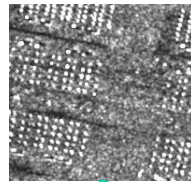


Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft



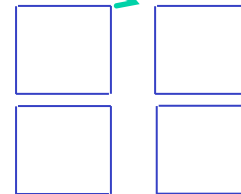
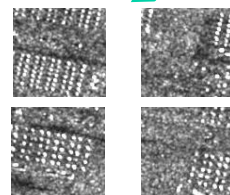
Cascaded Active Learning

Level 0
200x200 pixels

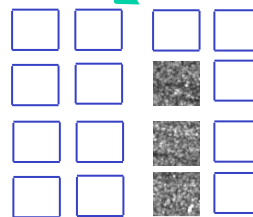
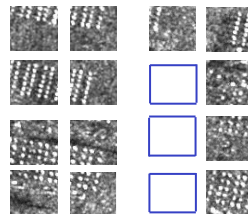


Refugee camp in Jordan

Level 1
100x100 pixels



Level 2
50x50 pixels



*Semantic
category*

Tents

Sand

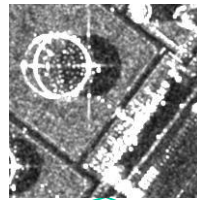


Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft



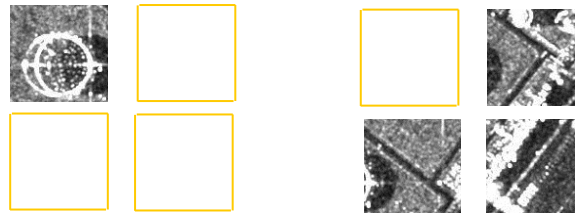
Cascaded Active Learning

Level 0
200x200 pixels



Petroleum storage area near Riffa, Bahrain

Level 1
100x100 pixels



Level 2
50x50 pixels

...

...

*Semantic
category*

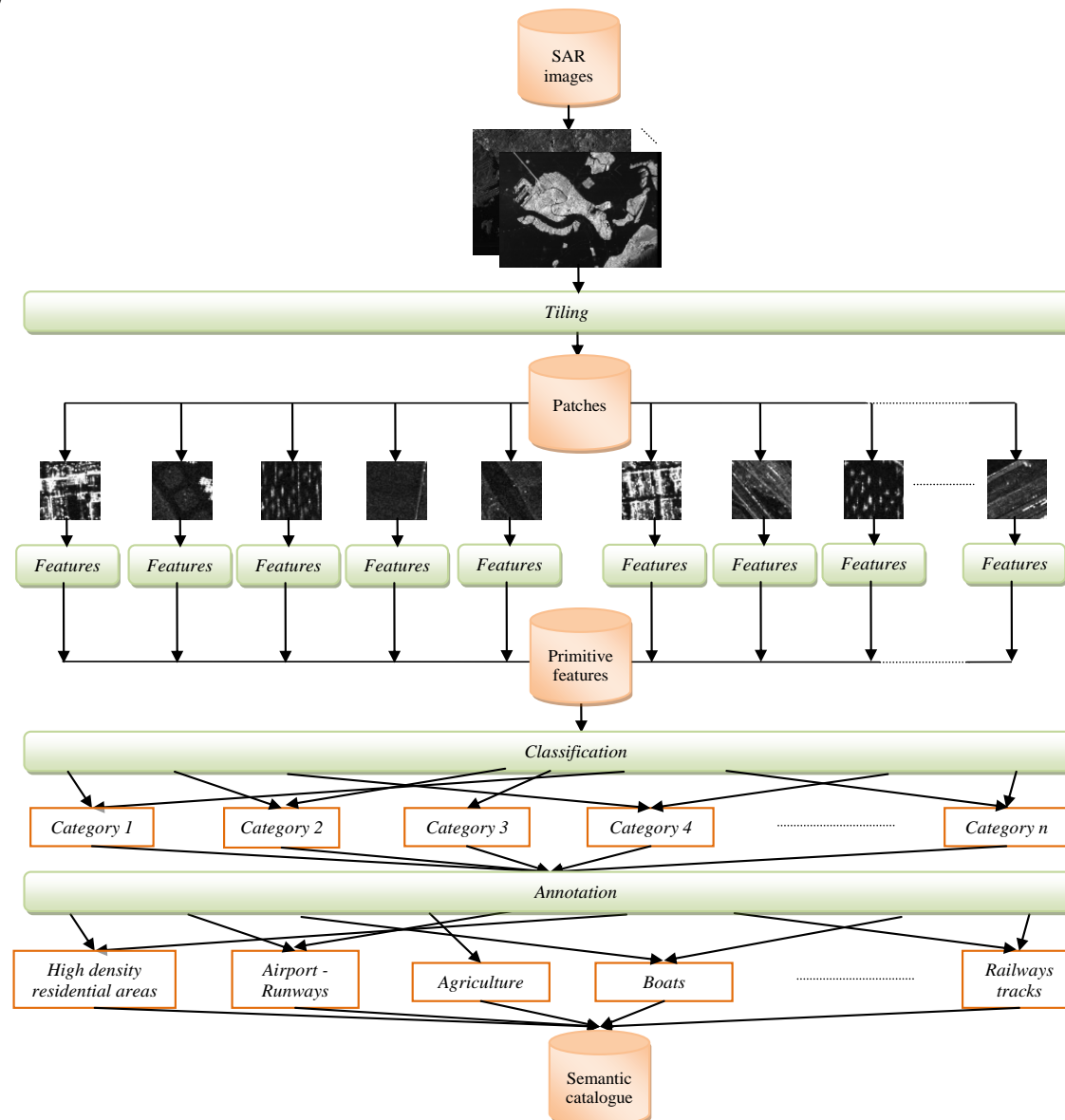
Storage tanks

Industrial buildings



Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

Ontology





Content Semantic Annotation

➤ Proposed three-level annotation scheme

Settlements

- Inhabited built-up areas
 - High density residential areas
-
- Uninhabited built-up areas
 - Skyscrapers



Industrial production areas

- Industrial facilities
 - Industrial buildings
-
- Industrial storage areas
 - Depots and dumps



Military facilities

- Air force facilities



Agriculture

- Greenhouses



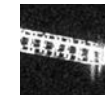
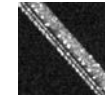
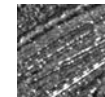
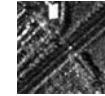
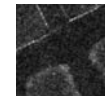
Natural vegetation

- Mixed forest



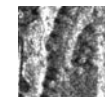
Transport

- Airports
 - Runways
-
- Roads
 - Streets and roads
-
- Railways
 - Railway tracks
-
- Bridges and tunnels
 - Bridges and fly-overs
-
- Ports and shipbuilding facilities
 - Harbour infrastructure
-
- Water vessels
 - Small vessels (boats)



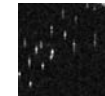
Bare ground

- Mountain



Water bodies

- Buoys



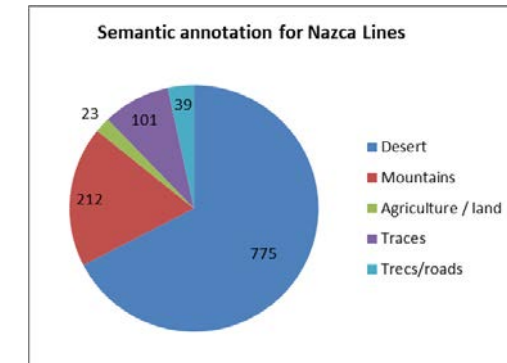
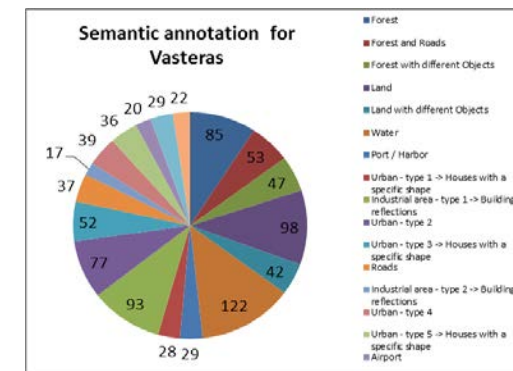
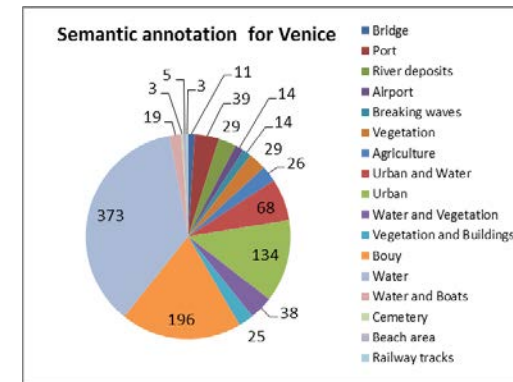


Semantic catalogues



350 cities
850 classes

-Bangkok (Thailand);
-Shenyang (China);
-Nazca Lines (Peru);
-Havana (Cuba);
-Venice (Italy);
-Vasteras (Sweden);
-Oran (Algeria);
-Bogota (Columbia) ...





Results: Venice

- **Investigated area** - Venice, Italy
- **Validation** – our results were compared for the same area with CORINE Land Cover (CLC) categories.



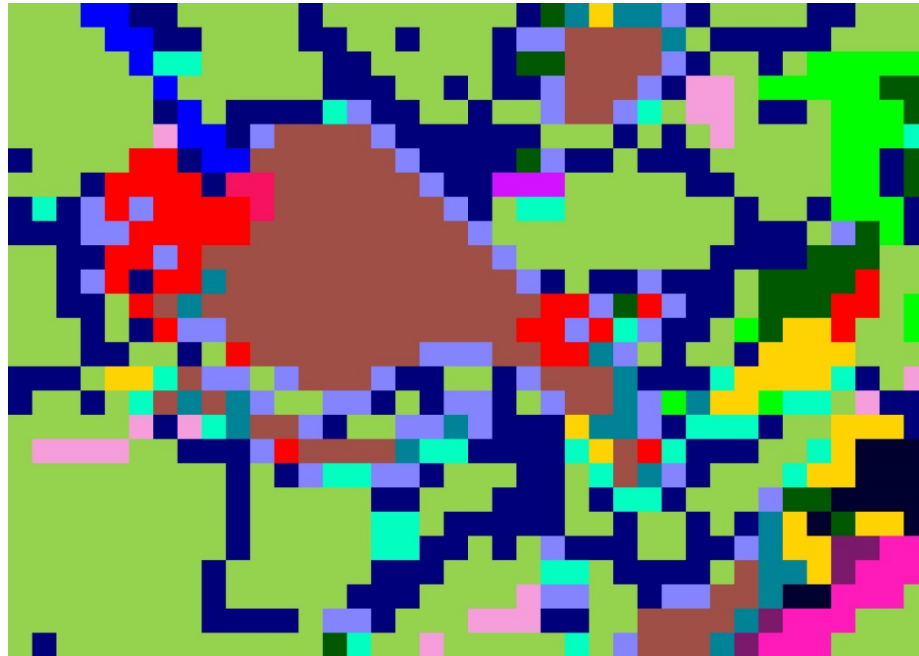
TSX image

- **Further validation** – our results to be compared with Urban Atlas categories.





Results: Venice

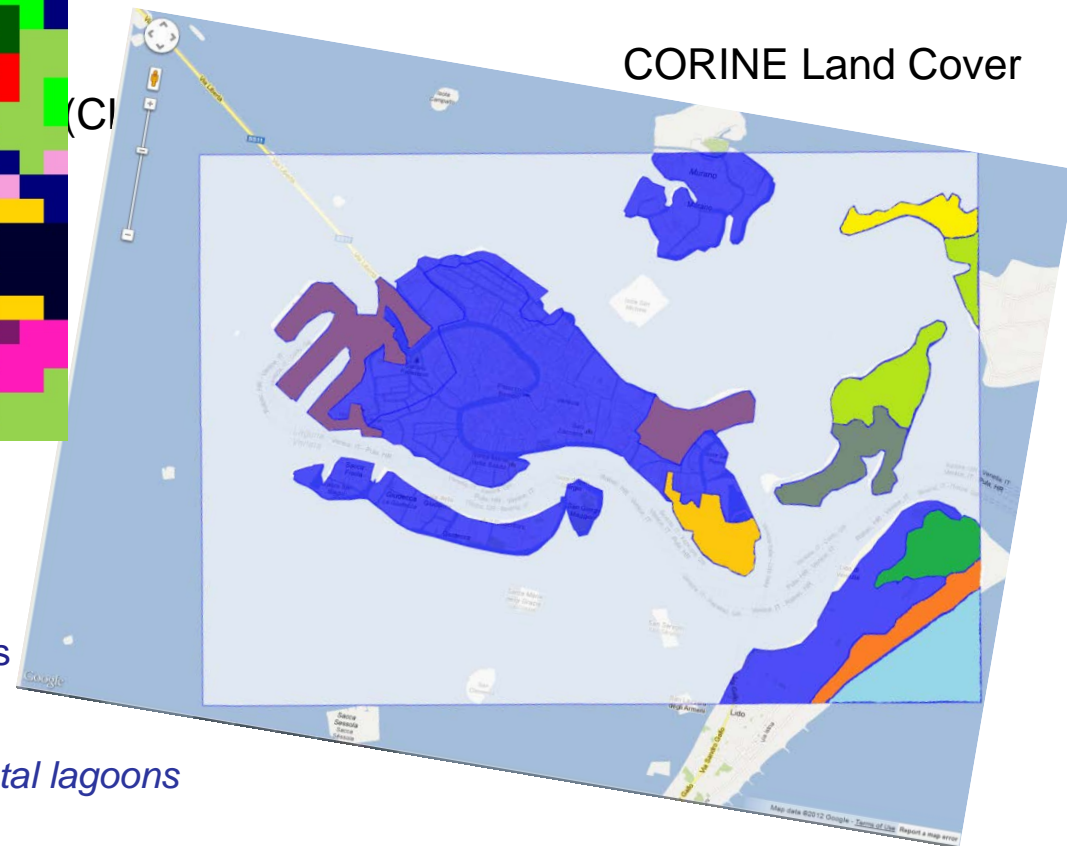


our classification

!!! *bridges, buoys, and sea categories*
of our proposed annotation method
are included in *marine waters – coastal lagoons*

in the case of CLC

Our classification – 17 categories
versus **CLC** – 10 categories

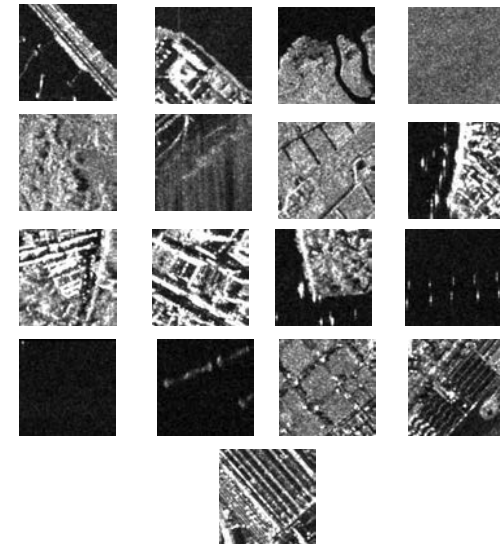
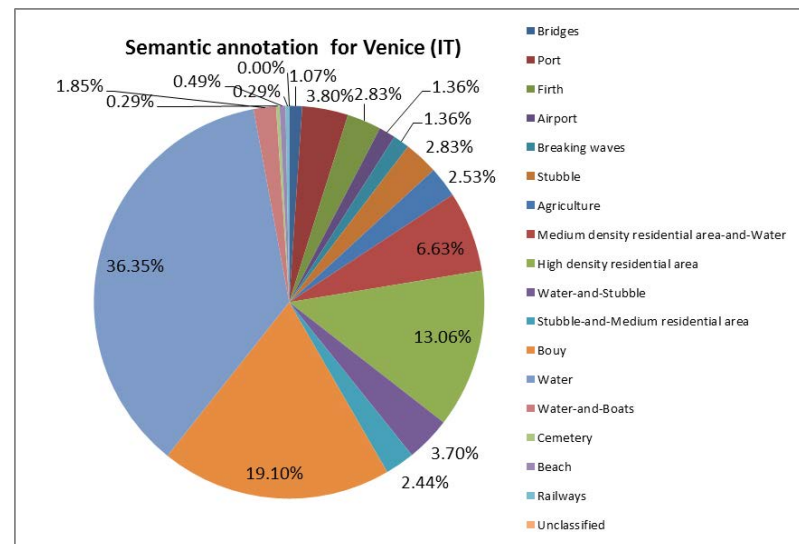




Results: Venice

➤ Data analysis

Percentage of patches per semantic category for Venice and a typical patch per category

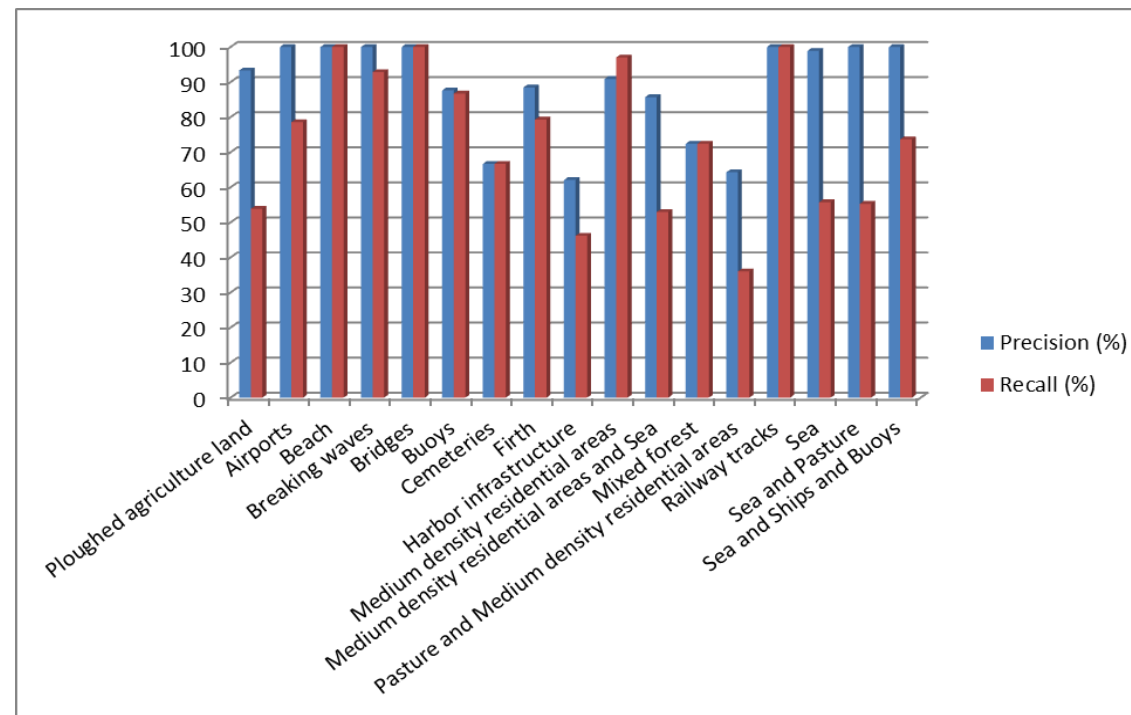




Results: Venice

➤ Quantitative results

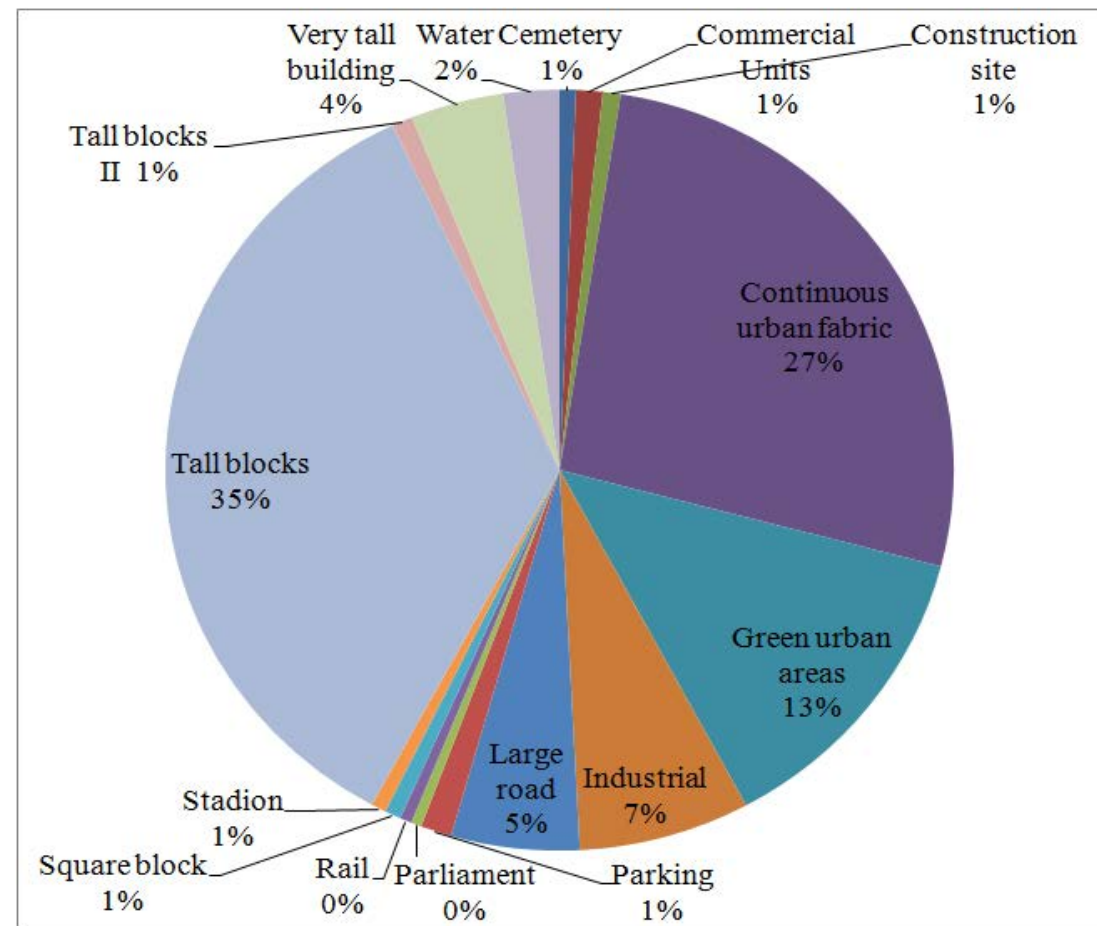
Precision / recall per semantic category for Venice





SCENE CATEGORIES & INFORMATION CONTENT: BUCHAREST

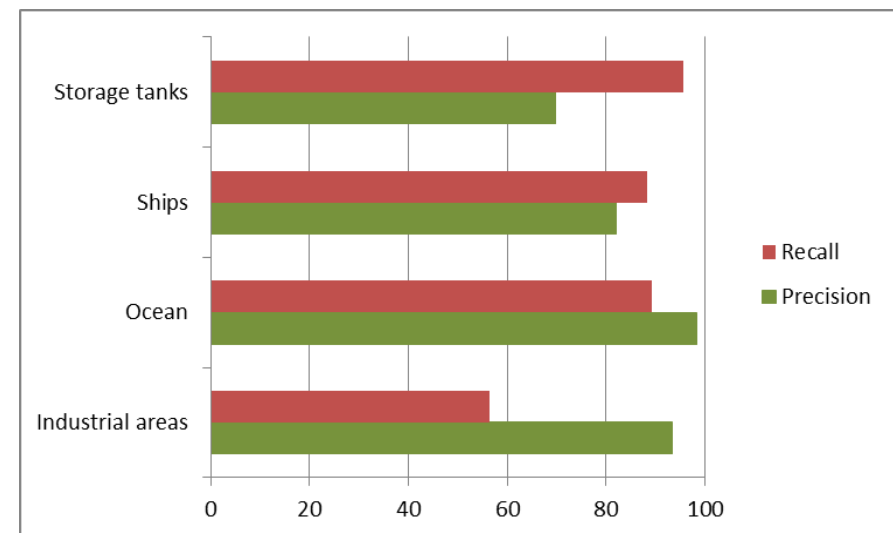
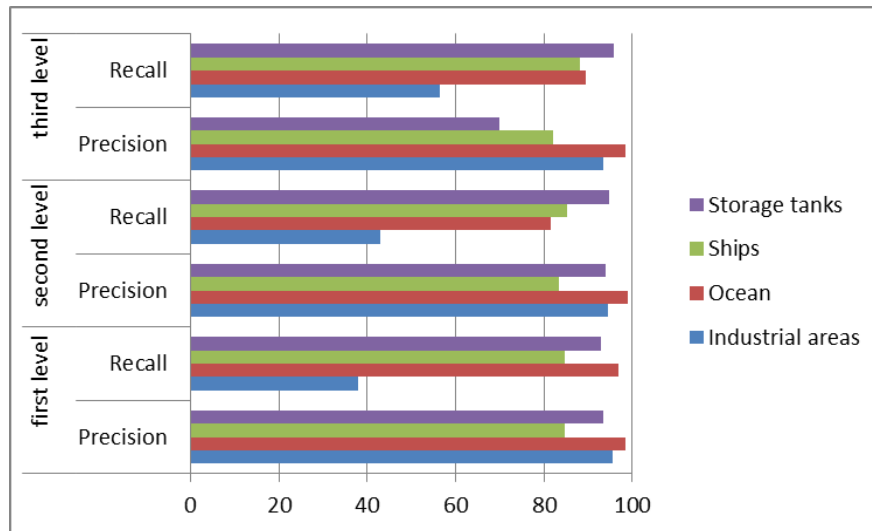
1 HS TerraSAR-X Scene =
up to 10 000 image patches
(100 x 100 m)





Evaluation

- **Evolution of precision/recall** results among four categories (*storage tanks, ships, ocean, and industrial areas*) out of seven for all levels (left side) and for the finest-level (right side).





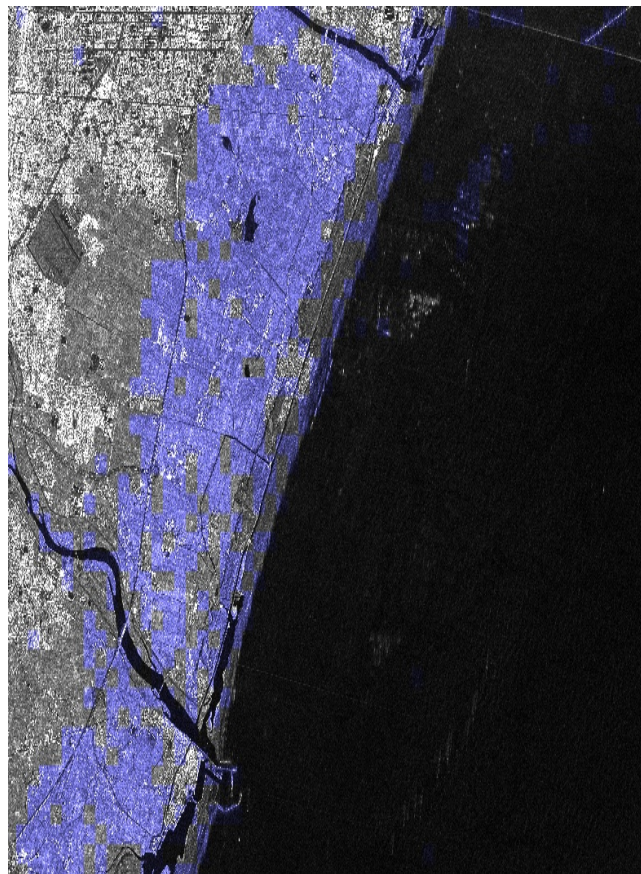


Disaster effects analysis

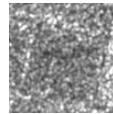
The damages in the agriculture can be clearly seen by comparing the classification in pre disaster image (left figure) with the post disaster image (right figure).

TerraSAR-X scene **before** Tsunami – 20.10.2010
Tsunami – 12.03.2011

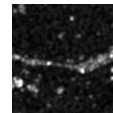
TerraSAR-X scene **after**



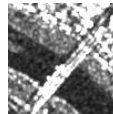
Agriculture



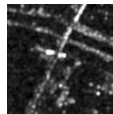
Flooded areas



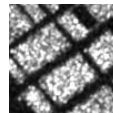
Bridges



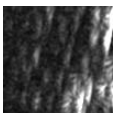
Bridges



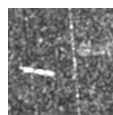
Aquaculture



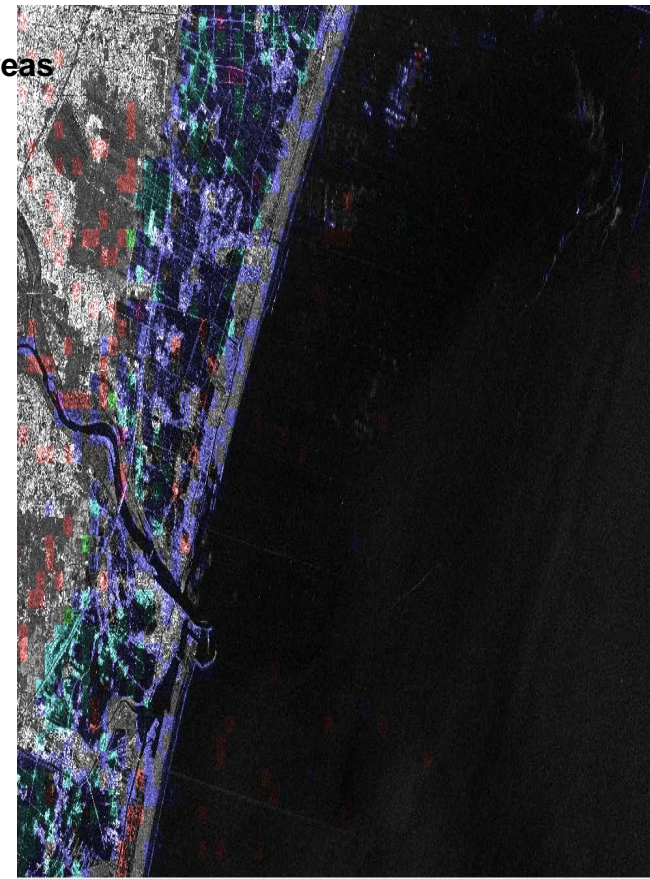
Debris



H. Voltage poles

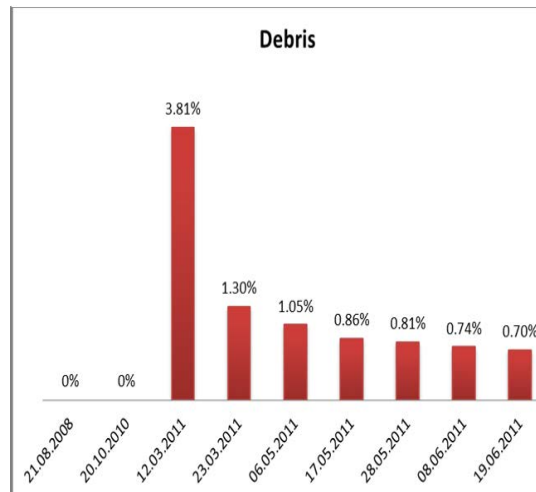
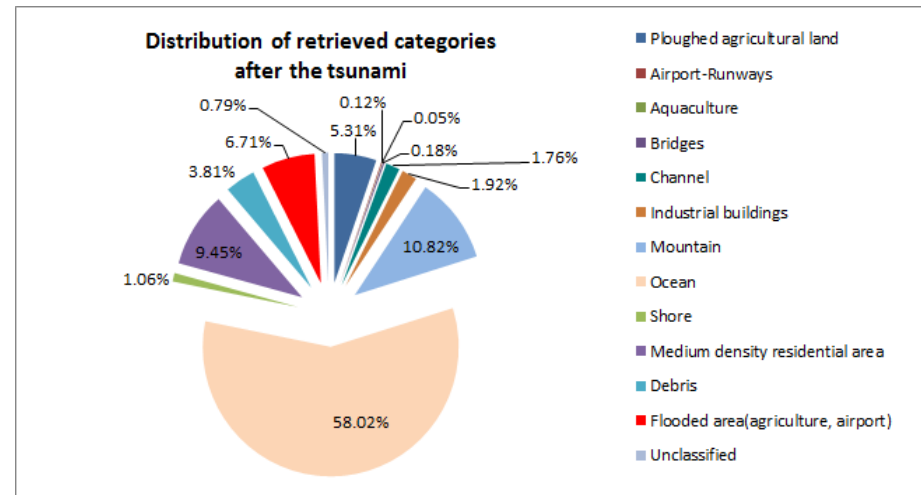
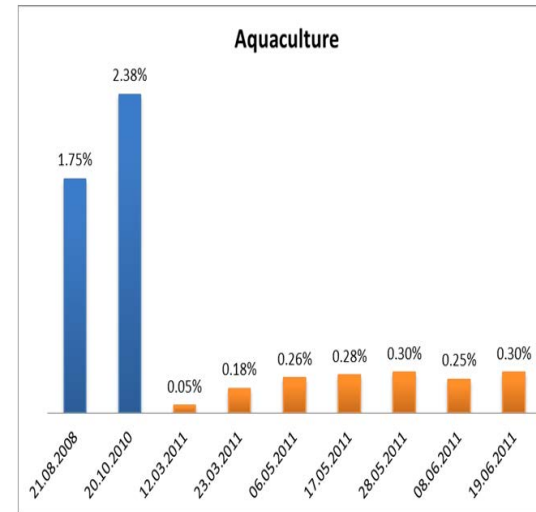
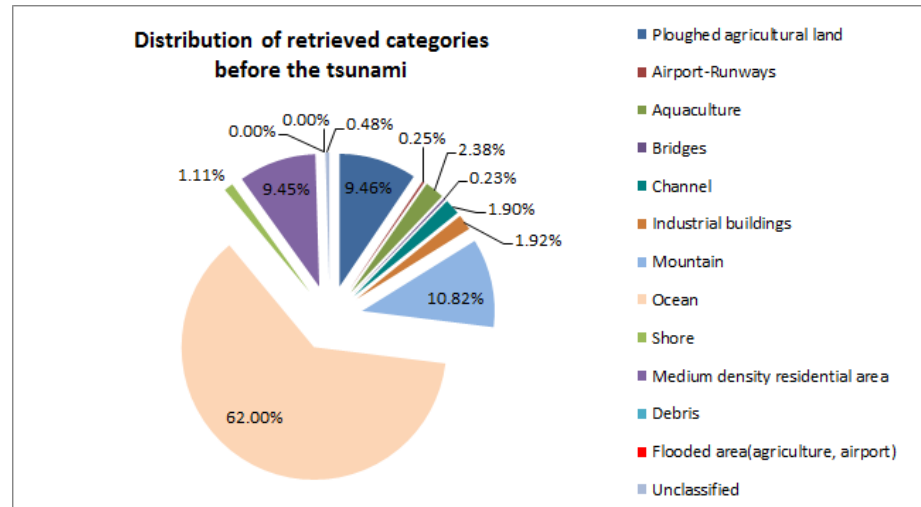


H. Voltage poles





Data Analytics





Conclusions

- Operational Data Mining, Visual Data Mining, KDD used to define semantic annotations of the image content.
- Interactive search supported by active learning mechanisms
- More than 1000 detailed categories of buildup scenes
- Time Series Exploration and Analysis
- Big Data



Acknowledgment:

The VDM component was developed by TERRASIGNA

References

- P. Blanchart, P., M. Ferecatu, C. Shiyong Cui, M. Datcu, 2015, Pattern Retrieval in Large Image Databases Using Multiscale Coarse-to-Fine Cascaded Active Learning, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume: 7, Issue: 4, Pages: 1127 – 1141.
- D. Molina Espinoza, M. Datcu, 2013, Earth-Observation Image Retrieval Based on Content, Semantics, and Metadata, IEEE Transactions on Geoscience and Remote Sensing, Vol. 51, No. 11, pp. 5145-5159.
- C. Dumitru, S. Cui, D. Faur, M. Datcu, 2014, Data Analytics for Rapid Mapping: Case Study of a Flooding Event in Germany and the Tsunami in Japan Using Very High Resolution SAR Images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,.
- O. Dumitru, M. Datcu, 2013, Information Content of Very High Resolution SAR Images: Study of Feature Extraction and Imaging Parameters. IEEE Transactions on Geoscience and Remote Sensing, Vol. 51, No. 8, pp. 4591-4610.
- M. Datcu, K. Seidel, 2005, Human Centered Concepts for Exploration and Understanding of Images, IEEE Trans. on Geoscience and Remote Sensing, ISSN 0196 2892, Vol. 43, No.3, pp. 601- 609.
- M. Datcu, H. Daschiel, et al, 2003, Information mining in Remote Sensing Image Archives: System Description, IEEE Trans. on Geoscience and Remote Sensing, ISSN 01 96 2892, Vol. 41, No. 12, pp. 2923 - 2936.